

**Method of rapid detection of mutations and nucleotide polymorphisms using chemometrics.**

**Field of Invention**

5

The present invention relates to the field of molecular biology and more specifically to methods and kits for detection of hybrid polynucleotides between a target and a labelled probe polynucleotide. The methods and kits may be used for the determination of mutations and polymorphisms in samples containing target  
10 polynucleotides. All patent and non-patent references cited in the application are hereby incorporated by reference in their entirety.

**Background of Invention**

15

Detection of hybridised double-stranded polynucleotides

Methods presently available for detecting nucleic acid sequences involve a relatively large number of steps and are laborious to perform. Some of the well-established methods are described for example in Hybridisation, by B. D. Hames and S. J. Higgins (Eds) IRL Press 1985 and J. A. Matthews et al, Analytical Biochemistry,  
20 169, 1988, 1-25 (Academic Press). Typically the DNA to be assayed is denatured and bound to a solid support, e.g. beads, nitrocellulose or nylon, and then incubated with a probe sequence, complementary to the target, which contains a label, for example  $^{32}\text{P}$ , or one half of an affinity pair, for example biotin. After incubation and  
25 washing of the solid phase, it is then developed to give a signal, by autoradiography for  $^{32}\text{P}$ , or in the case of biotin, typically by the addition of an avidin-enzyme conjugate, further washing and then substrate. Also sandwich assays are known, whereby a capture nucleic acid probe, of sequence complementary to the target, is bound to a convenient solid support. This is then allowed to hybridise to the target  
30 sequence. A labelled probe is then allowed to hybridise to another part of the target sequence, and after washing steps, the label is developed as appropriate. Alternatively, the "sandwich" is formed in solution with two probes, one containing the label and the other containing one half of an affinity pair. After hybridisation, the solution is then contacted with the solid phase to which the other half of the affinity  
35 pair is bound. Washing and signal development then takes place. All these methods

and their variants require numerous handling steps, incubations, washings, and the like which result in processes which are both laborious and time-consuming.

5 Nucleic acid sequences may be amplified prior to hybridisation to increase the number of target nucleic acids. Methods for detecting amplified nucleic acid sequences, for example DNA sequences generated by PCR, are as outlined above. Alternatively, they can be analysed by gel electrophoresis, and the existence of a band of a given molecular weight is taken as evidence of the presence of the sequence in the original DNA sample, as defined by the specific primers used in the  
10 amplification reaction. Again such methods are relatively time-consuming and laborious.

Many hybridisation assays include some sort of signal amplification, which may comprise binding of more labelled probe to the same target nucleic acid or the use  
15 of an enzyme based system, wherein the amount of signal produced per probe can be increased by letting the enzymes convert more substrate to detectable product. A further example of signal amplification involves the use of nucleotide amplifiers as described in US 5,635,352 and US 5,124,246 (assigned to Chiron).

20 Common for all the known methods is that it is always necessary to perform some sort of washing to remove any unhybridised probe, which will give rise to the same signal as hybridised probes. In many cases, it is necessary to perform a whole series of washing operations to reduce background or false positive signal. These hybridisation and washing steps need to be performed by a technician and/or a  
25 robot programmed to perform the routine steps thus contributing significantly to the cost of the assays.

Signal detection in most of the known methods is relatively simple in that only one signal is determined and this is done at one wavelength only. Most often, the  
30 determination is a qualitative determination of whether there is a significant signal or not. There are also methods available for the quantitative detection of hybridisation between target and probe. These methods may be used to determine the amount of target nucleotides in the sample but are likewise restricted to data detected at one wavelength.

35

Because signal detection is very simple it is generally not possible to distinguish a false positive signal from a true positive one. Thus, if an error is made in any of the hybridisation and/or washing steps then a false positive may be obtained and there is no way of telling this from a true positive one.

5

Accordingly, there is a need for development of simple hybridisation assays without the need for washing steps and for development of methods allowing distinction between the signal from an unhybridised probe from the signal from a hybridised probe.

10

### **Mutations**

15

With the advent of the modern molecular genetics, it is becoming increasingly clear that a substantial part of human diseases have a genetic component. Whereas the initial discoveries within the field of medical genetics were dominated by relatively rare diseases characterised by a strong one-mutation one-disease relation, emphasis is now being put on the investigation of frequent diseases such as diabetes, dyslipidemia, hypertension and cardiovascular diseases, wherein the one-mutation:one-disease relation does not always exist.

20

Most of these common diseases are polygenic, involving several loci, and many population association studies leave little doubt that the genotype of a number of the involved loci must be established in order to provide reliable data for use in diagnosis, prognosis and pharmacogenetics.

25

Similar considerations apply to other organisms apart from human. Several species are used for general genetic research as well as models for studying diseases. In addition, a lot of genetic research is performed on livestock, plants, infectious agents etc. for various purposes like increasing yields in agricultural areas and preventing diseases in livestock, in commercial plants, and in humans.

30

35

Polymorphisms and mutations are ubiquitous in the genome. The distinction between polymorphisms and mutations lies in the frequency of alleles in a gene, where a mutation is defined if one allele amounts to more than 99% of the alleles, and polymorphisms accounts for all other situations. Both polymorphisms and

mutations can be single nucleotide substitutions, deletions, insertions or rearrangements. As a consequence of the Human Genome Project interest has been focussed particularly on single nucleotide polymorphisms (SNP) for various reasons including their usage in genome scanning for diseases-related genes. More than 2 million SNP's has been detected, some but not all of which are directly related to diseases.

Often more than one polymorphism, in particular SNP, are present in the same gene. The significance of this varies, but in coding regions, i.e. regions of the gene coding for the proteins, this is of particular interest: in diploid organisms as human it is of importance on which of the two alleles the polymorphisms are located, as this will determine if 0, 1 or several changes in the same protein is present and consequently influence the function of the protein differently.

A number of methods exist that can be used to detect SNPs, mutations and other polymorphisms. Detailed description of useful methods may be found in Ausubel et al. Current protocols in molecular biology, (2000) John Wiley and Sons, Inc., N.Y. and in Sambrook et al., Molecular Cloning, A Laboratory Manual, 2nd Edition (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Among the more important methods are: DNA sequencing (Sanger et al. (1977) Nature, 265: 678-695; Maxam and Gilbert (1977) Proc Natl Acad Sci USA, 74: 560-564); the single strand conformation polymorphism (SSCP) method (Orita et al. (1989) Genomics, 5:874-879); the denaturing gradient gel electrophoresis (DGGE) method (Fisher and Lerman. (1993) Proc. Natl. Acad. Sci. U.S.A., 80:1579-1583) and later improvements of the SSCE and the DGGE techniques, such as the dideoxy fingerprinting (Sarkar et al., Genomics, 13:441-443 (1992), restriction endonuclease fingerprinting (Liu and Sommer, Biotechniques, 18:470-477 (1995)) and constant denaturing gel electrophoresis (CDGE), (Hovig et al., Mut. Res., 262:63-71 (1991)). However, all of these methods are labour-intensive, and involves fairly complex handling.

Mutations can be detected by specific hybridisation of oligonucleotides to the nucleic acid to be analysed.

US 6,022,686 (Zeneca Ltd) describes the simultaneous use of fluorescent probes for the detection of mutations by liquid phase hybridisation and fluorescence polarisation. Since the detection occurs by hybridisation of nucleic acid sequences in homogeneous solution, the method allows detection of probe specific sequences from both chromosomes.

Variations in DNA-sequences may also be detected by hybridisation in a system consisting of a solid and a liquid phase (two-phase systems). One particular useful approach is the so-called "sandwich" hybridisation described in US 4,486,539 (Orion Corp. Ltd.) . Many haplotypes are described by single base polymorphisms, but US 4,486,539 do not provide methods to detect single base polymorphisms. The same is true for the majority of the methods for detection of hybrid nucleotides described above. For most of these methods, it is very difficult to design conditions under which one probe is capable of distinguishing between target nucleotides differing in one position only.

This is however possible with the method described in US 5,633,134 (IG Laboratories) wherein mutations are detected by hybridisation in the presence of a quaternary ammonium salt is described. This reference also devise methods for detection of multiple mutations in one gene but only by hybridisation in the presence of a quaternary ammonium salt

In the last several years there have been put a lot of effort in developing techniques for simultaneous detection of multiple mutations or polymorphisms.

Thus, the existent methods for determining mutations and/or polymorphisms have significant limitations and further improvements are greatly needed for their use in large-scale genetic screening of clinical material.

### **Definitions**

Recording spectral data: a spectrum is understood as a spectrum covering a range of wavelengths of electromagnetic radiation, including vacuum ultraviolet, ultraviolet, visible, near infrared, infrared, far infrared light, microwaves, and radiowaves. The

spectral data must be in a format so that they can be exposed to multivariate analysis.

5 For the purposes of the present invention, performance of mass spectrometry is also regarded as recording a spectrum, although in this case no electromagnetic radiation is emitted or recorded. Also in this case no detectable label is used.

10 Label: for the purpose of the present invention a detectable label refers to a chemical moiety capable of emitting, absorbing or scattering electromagnetic radiation. More preferably, a label refers to chemical moieties capable of emitting, absorbing, or scattering phosphorescence, luminescence, or fluorescence.

15 Hybrid polynucleotide: a hybrid between two polynucleotides, wherein the association between the two polynucleotides is stronger than the association between one of the polynucleotides and water.

### Summary of invention

20 Accordingly, the methods according to the present invention provide simple analytical techniques compared to the prior art methods for determination of hybrid polynucleotides. This will render the analytical techniques according the present invention available for laboratories without special molecular biological expertise. The methods are simple and cheap both with regard to equipment and use.

25 The invention will also solve some of the problems of the research into diseases having a genetic factor. The problem is that many of the important diseases, such as diabetes, artherosclerotic and psychiatric diseases are of polygenic nature meaning that the interaction between several genes leads to a predisposition or directly leads to these diseases. Elucidation of the genetic component of these diseases requires  
30 a large number of patients and a significant analytical effort. The present invention partly reduces the costs associated with these analyses and partly accelerates the investigations.

It is expected that the present invention will partly replace the main part of the gene-diagnostic methods based on electrophoresis and sequencing. It is likewise expected to supplement the so-called chip based gene-diagnostic methods.

5 In a first main aspect, the invention relates to a method for establishing whether at least one target polynucleotide is present in a sample, comprising the steps of

- 10 i) providing a sample to be analysed for the presence of the at least one target polynucleotide,
- ii) adding at least one polynucleotide probe at least partly complementary to a sub-sequence of the at least one target polynucleotide, wherein the at least one probe comprises at least one detectable label,
- 15 iii) incubating the sample under conditions suitable for the formation of at least one hybrid polynucleotide comprising the at least one probe and the at least one target polynucleotide, when present,
- 20 iv) recording spectral data from an environment comprising at least part of the sample,
- v) analysing the spectral data, and
- 25 vi) establishing whether the target polynucleotide is present.

When the target polynucleotide is present, the at least partly complementary probe is capable of hybridising to the sub-sequence of the target polynucleotide.

30 In one embodiment of this method of the invention, the analysis of the spectral data can distinguish for each of the at least one probe whether the probe is part of the at least one hybrid polynucleotide or not part of the at least one hybrid polynucleotide.

In a further embodiment of this method of the invention, the analysis of the spectral data can distinguish for each of the at least one probe, when the probe is part of the

at least one hybrid polynucleotide, whether or not there is a mismatch between the probe and the sub-sequence of the at least one target polynucleotide.

5 In preferred embodiments of the method, the spectral data are analysed using multivariate analysis.

10 In a further aspect the invention relates to a method for identifying a hybrid polynucleotide comprising at least partly complementary nucleotide strands, said method comprising the steps of

providing a sample comprising at least one target polynucleotide,

15 providing at least one polynucleotide probe at least partly complementary to the at least one target polynucleotide,

wherein the polynucleotide probe comprises at least one detectable label,

20 forming a hybrid polynucleotide comprising at least one target polynucleotide and at least one polynucleotide probe, and

recording spectral data from an environment comprising the hybrid polynucleotide, and

25 wherein the spectral data when at least one oligonucleotide probe forms part of the hybrid polynucleotide are different from the spectral data when at least one oligonucleotide probe is not part of the hybrid polynucleotide.

30 By using spectral data instead of only data from one wavelength it is possible to determine the presence or absence of a particular hybrid polynucleotide even in the presence of the unbound polynucleotide probe. It is thus possible to record a spectrum from an environment comprising both the hybrid and unbound probe and by comparing that spectrum with spectra of the unbound probe and the hybrid determine whether a hybrid has formed or not.



It is also possible to distinguish the spectrum from a hybrid between target and probe with a given number of mismatches from the spectrum of a hybrid between target and probe with a higher number of mismatches. It is thus e.g. possible to distinguish the spectrum from a hybrid where there is a stretch of 100% complementarity between probe and target (i.e. zero mismatches) from a spectrum recorded from a hybrid, where there is just one mismatch between target and probe. It is also possible to distinguish the spectrum from a hybrid where there is one mismatch between probe and target from a spectrum recorded from a hybrid, where there is more than one mismatch between target and probe.

The analytical method is fast, easy to use and can be completely automated. It is easy to include various control steps in the method and thereby ensure that the detected results can be verified.

In a further aspect the invention relates to a method for detecting a hybrid between a target polynucleotide and a probe oligonucleotide comprising

providing a sample comprising at least one target polynucleotide,

providing at least one polynucleotide probe at least partly complementary to the target polynucleotide,

wherein the polynucleotide probe comprises at least one detectable label,

forming a hybrid polynucleotide comprising at least one target polynucleotide and at least one polynucleotide probe,

recording spectral data representing electromagnetic radiation, said spectral data being recorded from an environment comprising the hybrid polynucleotide, and

treating the spectral data using multivariate analysis to correlate the recorded data to the presence or absence of at least one hybrid.

By using multivariate analysis of spectral data, more information can be gathered from one single hybridisation procedure. Moreover the analysis allows extraction of

information which is otherwise "hidden" when looking at the raw data. For example as described in the detailed description part of the present invention, it is possible to group spectra from different signal sources (probe alone, probe-target with 100 % complementarity, probe-target with mismatch, two probes- one target, etc) and in this way place an unknown spectrum in the correct group. This allows the use of two or more hybridisation reactions to be carried out at the same time, because the data originating from each hybrid can be distinguished by the multivariate analysis. The method also allows determination of the presence or absence of a hybrid without washing away unbound probe, because the multivariate analysis can distinguish the two after appropriate calibration.

In a further aspect the invention relates to a method for detecting a mutation or polymorphism, said method comprising the steps of

providing a sample comprising at least one target polynucleotide comprising a preselected region suspected of containing the mutation and/or polymorphism,

providing at least one polynucleotide probe capable of hybridising specifically with at least one individually selected nucleotide sequence in said preselected region,

wherein the polynucleotide probe comprises at least one detectable label,

contacting the target and probe under conditions allowing the formation of a hybrid polynucleotide comprising at least one target polynucleotide and at least one polynucleotide probe,

recording spectral data representing electromagnetic radiation, said spectral data being recorded from an environment comprising the hybrid polynucleotide, and

correlating the spectral data using multivariate analysis to the presence or absence of said mutation or polymorphism.

By this method determination of a mutation can be performed easily without the use of demanding purification steps. By the method according to the invention it is possible to distinguish among signal from unhybridised probe, probe-wildtype, and

probe-mutant. The method also allows the determination of more than one mutation at one time in particular the determination of whether two or more mutations/polymorphisms are located on the same chromosome.

5 In a further aspect the invention relates to a kit for detection of mutations or polymorphism comprising

at least one oligonucleotide probe capable of hybridising to a preselected region of a target polynucleotide, the polynucleotide probe further comprising at least one  
10 detectable label,

instructions enabling correlation of spectral data recorded from a hybrid polynucleotide between said at least one oligonucleotide probe and said target polynucleotide to the presence or absence of said mutation or polymorphism using  
15 multivariate analysis.

The instructions may be in the form of calibration data relating to the specific assay to which the kit relates, so that the user of the kit does not have to carry out extensive analyses to determine the grouping of spectra in multivariate analysis.  
20

In a further aspect, the invention relates to a system for establishing whether at least one target polynucleotide is present in a sample, comprising

- i) at least one polynucleotide probe being at least complementary to a target polynucleotide, the probe comprising a detectable label,
- 25 ii) a sample chamber from which electromagnetic radiation can be recorded,
- iii) a source of spectrally resolved electromagnetic radiation,
- iv) means for sensing and recording a spectrum of electromagnetic radiation from the sample chamber, and
- 30 v) a computer unit for storing spectral data of electromagnetic radiation and having instructions to treat the recorded spectral data using multivariate analysis.

In an even further aspect, the invention relates to a system for detection of a hybrid  
35 polynucleotide comprising

- i) at least one oligonucleotide probe being at least partly complementary to a target polynucleotide, the probe comprising a detectable label,
- ii) a sample chamber from which electromagnetic radiation can be recorded,
- 5 iii) a source of spectrally resolved electromagnetic radiation,
- iv) means for sensing and recording a spectrum of electromagnetic radiation from the sample chamber, and
- v) a computer unit for storing spectral data of electromagnetic radiation and having instructions to treat the recorded spectral data using
- 10 multivariate analysis.

The system is adapted for performance of the methods according to the present invention and provides a system which allows high throughput screening of

15 samples.

### Description of Drawings

**Figure 1.** The underlying idea of PCA modelling is to replace a complex multidimensional data set by a simpler version involving fewer dimensions, but still fitting the original data closely enough to be considered a good approximation. A 3-D data swarm is shown. The new axes (the principal components PC1 and PC2) are placed in the directions of the largest variances. In this example 3 dimensions are reduced to 2.

20

**Figure 2.** A PC1/PC2 scores-plot. PC 1 is the first principal component, which is most important in explaining the difference between the samples. PC 2 is the next important, PC 3 the third and so on.

25

**Figure 3.** A PC1/PC2 loading-plot.

30

**Figure 4.** Nucleotide section of the Apo3611 region. Complementary wildtype and mutant oligonucleotide probe.

**Figure 5.** PCA scores plot based on emission spectra obtained from the room temperature hybridisations of Cy5 labelled oligonucleotides (A: Cy5 labelled

35

wildtype oligonucleotide interrogating the wildtype nucleotide at the 5'end, **B**: Cy5 labelled mutant oligonucleotide interrogating the wildtype nucleotide at the 5'end, **C**: Cy5 labelled wildtype oligonucleotide interrogating the wild type nucleotide in the central part of the probe or **D**: Cy5 labelled mutant oligonucleotide interrogating the wildtype nucleotide in the central part of the probe) in the presence or absence of wildtype target DNA (**W1**). The spectra were either obtained from the oligonucleotides alone or in combination with target (denoted X\_W). The first prefix (1 or 2) gives the replicate number, whereas the second prefix represent two fluorescence measurements on that same replicate solution. PCA scores plot from experiment 1.

**Figure 6.** PCA scores plot based on emission spectra obtained from 30°C hybridisations of Cy5 labelled oligonucleotides (A, B, C or D) in the presence or absence of wildtype target DNA (**W1**). The prefix (1-3) gives the replicate number. PCA scores plot from experiment 2.

**Figure 7.** PCA scores plot based on emission spectra obtained from 30°C hybridisations of Cy5 labelled oligonucleotides (A, B, C or D) in the presence of wildtype target DNA (**W1**). Data points are the same as for Figure 6 but spectra of fluorescent oligonucleotides without targets have been left out. The prefix (1-3) gives the replicate number. PCA scores plot from experiment 2.

**Figure 8.** PCA scores plot based on excitation spectra obtained from 30 °C hybridisations of Cy5 labelled oligonucleotides (A, B, C or D) in the presence or absence of wildtype target DNA (**W1**). The prefix (1-3) gives the replicate number. PCA scores plot from experiment 2.

**Figure 9.** PCA scores plot based on excitation spectra obtained from 30 °C hybridisations of Cy5 labelled oligonucleotides (A, B, C or D) in the presence of wildtype target DNA (**W1**). Data points are the same as for Figure 8 but spectra of fluorescent oligonucleotides without targets have been left out. The prefix (1-3) gives the replicate number. PCA score plot from experiment 2.

**Figure 10.** PCA scores plot based on excitation spectra obtained from 30 °C hybridisations of Cy5 labelled oligonucleotide A in the presence or absence of

wildtype target DNA (W2), mutant target (M) or non-specific target (Z). The prefix (1-3) gives the replicate number. PCA scores plot from experiment 3.

5 **Figure 11.** PCA scores plot based on emission spectra obtained from 30 °C hybridisations of Cy5 labelled oligonucleotide A in the presence or absence of wildtype target DNA (W2), mutant target (M) or non-specific target (Z). The prefix (1-3) gives the replicate number. PCA scores plot from experiment 3.

10 **Figure 12.** PCA scores plot based on excitation spectra obtained from 30 °C hybridisations of Cy5 labelled oligonucleotide A in the presence or absence of wildtype target DNA (A-W2), mutant target (A-M) or both DNA targets simultaneously (A-W2-M, note concentrations: 0,125 or 0,25 µM). The measurement A-W2-M-2\_2 is not included while it was considered as an outlier. The prefix (-1 or -2) represent two different concentrations of the targets, and the prefix (1-3) give the replicate number. Data are from experiment 4. An outlier has been removed from the data analysis.

20 **Figure 13.** PCA scores plot based on emission spectra obtained from 30 °C hybridisations of Cy5 labelled oligonucleotide A in the presence or absence of wildtype target DNA (A-W2), mutant target (A-M) or both DNA targets simultaneously (A-W2-M). The prefix (-1 or -2) represent two different concentrations of the targets, and the prefix (1-3) give the replicate number. Data are from experiment 4.

25 **Figure 14.** PCA scores plot based on Cy5 excitation spectra obtained from 30 °C hybridisations of Cy5 labelled oligonucleotides A or B and with or without Cy3 labelled oligonucleotide F in the presence or absence of wildtype target W3. The prefix (1-3) gives the replicate number. Data are from experiment 5.

30 **Figure 15.** PCA scores plot based on emission spectra obtained from 30 °C hybridisations of Cy5 labelled oligonucleotides A or B and with or without Cy3 labelled oligonucleotide F in the presence or absence of wildtype target W3. The prefix (1-3) gives the replicate number. Data are from experiment 5.

**Figure 16.** Illustrates fluorescence spectra of a target polynucleotide (W1 of Figure 4) alone and the PCR buffer used for amplification alone.

5 **Figure 17.** Shows on the left fluorescence spectra of two different probes labelled with Cy5 (A and B from Figure 4). On the right is shown the two spectra recorded from samples with both the probe and the target polynucleotide. The interaction between target and probe changes the spectrum.

10 **Figure 18.** Shows on the left fluorescence spectra of two different probes labelled with Cy5 (C and D from Figure 4). On the right is shown the two spectra recorded from samples with both the probe and the target polynucleotide. The interaction between target and probe changes the spectrum.

15 **Figure 19.** PCA-plot of dsDNA after hybridization to Cy5-labelled probe. In this analyses ssDNA-probe complexes are included for illustrative purposes (AM and AW2), but are usually not included in the analysis. AM-Rm, probe complexed with dsDNA mutant; AW2-R, probe complexed with dsDNA wildtype; AW2-R-, probe complexed with dsDNA wildtype, but the denaturation step is omitted; A, probe alone.

20 **Figure 20.** PCA-plot of Poly2-dsDNA after hybridization to Cy5-labelled probe. In this analyses ssDNA-probe complexes are included for illustrative purposes (dWP-WA and dWP-MA), but are usually not included in the analysis. dWP-WA-WS, probe complexed with dsDNA wildtype Poly2; dWP-MA-MS, probe complexed with dsDNA mutant Poly2.

25

## Detailed description of the invention

### Recording spectral data

30 In the following one example of recording of spectral data is illustrated with reference to Figures 16-18. These figures represent data from the experiments carried out in Experiment 1 of the examples. Figure 16 shows a fluorescence spectrum of the target DNA alone (top) and the PCR buffer alone (bottom). The differently shaded curves represent different replicate samples. In Figure 17 on the

35

left is shown spectra of the probes (A and B as described in the examples and Figure 4). Note the difference in the scale of the Y axis between Figure 16 and Figure 17/18. The buffer and target alone do not contribute significantly to the spectrum of Figures 17/18.

5

In probes A and B, the mutant nucleotide is placed terminally and the label is linked directly to the mutant nucleotide. On the right is shown similar spectra of samples with both the target polynucleotide (W1 as defined in Figure 4) and the two probes (A-wildtype and B-mutant). Probe A which is 100 % complementary to a sequence in the target polynucleotide interacts differently with this than probe B, which differs in one position. Through the interaction between target and probe, the recorded spectrum is changed slightly. In this particular case the difference is just visible in the spectrum, but in order to make full use of the information in the spectra multivariate analysis is used. These changes can be analysed using multivariate analysis. It can also be seen that the spectrum of probe B, which differs from probe A at one position (see Figure 4) is not changed to the same degree by interaction with the target polynucleotide.

In Figure 18 a further example is illustrated. The probes (C and D) differ from probes A and B in that the mutant nucleotide is placed centrally in the probe (see Figure 4). The spectra to the right of Figure 18 are recorded from samples with both the probe and the target polynucleotide, which is 100 % complementary to probe C and differs from probe D at one (terminal) position. Again the association between target and probe changes the spectrum so that the presence of a target wildtype can be distinguished from a target mutant polynucleotide.

### Data analysis

The art of extracting chemically relevant information from data produced in chemical experiments is given the name "Chemometrics" in analogy with biometrics and econometrics. Chemometrics is heavily dependent on the use of different kinds of mathematical models - high information models, ad-hoc models, and analogy models. The task demands knowledge of statistics, numerical analysis, operation analysis, etc., and applied mathematics. However, as in all applied branches of science, the difficult and interesting problems are defined by the applications.



There is a tendency to make fewer experiments and measure more and more data in each of them. This trend is seen everywhere, ranging from physics with high costs for accelerators and other equipment to biomedical research where ethical and regulatory aspects provide a strong incentive for fewer experiments.

Multivariate data analysis/chemometrics convert any more or less structured and complicated sets of data into a few meaningful plots that display the information hidden in the data structures in a way that is easy to understand.

10

### **Principal Component Analysis (PCA)**

Large data tables often contain a vast amount of partially concealed information, which is too complex to be readily interpreted. Principal Component Analysis (PCA) is a method of projection, which helps to visualise such information. PCA is instrumental in determining the manner in which a sample differs from another, the variables that contribute most to this difference and whether some of the variables contribute in the same way (are correlated) or whether they are independent of each other. PCA also gives the opportunity of finding patterns (clusters) in the data.

20

The principle of the analysis is to arrive at a simplified set of data (data reduction) whilst retaining as much information as possible. This means that an attempt is made at trying to determine which variables are essential to the total variation in the set of data and to determine whether some of the variables provide the same information (i.e. separate the samples in the same way). By doing this, an attempt is made at uncovering the underlying structure of the data material.

25

The principal components (PCs) are extracted one after the other by mapping where the greatest total variation in the data set is to be found. The first principal component is placed where the greatest distances between the samples are to be found, and the analysis provides a percentage value of how much of the total variation is explained by looking in the direction of the first principal component. The next principal components are found as directions/axes on which the rest of the variation in the set of data can be found. Gradually, as more and more principal components are found one will acquire more and more descriptions of the

30

35

differences between the samples with respect to the given variables. Usually, the data set includes noise and since this noise is not used to describe the differences between samples there is a limit to how much variation one can describe.

5 Figure 1 visualises how the 2 first principal components (PC1 and PC2) are found in a 3 dimensional plot. Only data of 3 variables are available, so the linear behaviour is immediately recognised by plotting the objects in the 3-D variable space. When more variables are present (like in spectroscopy where each row of a data matrix is a spectrum of perhaps several hundred of wavelengths), this procedure is of course  
10 not feasible. Identification of this type of linear behaviour in a space with several hundred dimensions of course cannot be achieved by visual inspection. Here PCA, with its powerful projection characteristics, helps to discover the hidden data structures.

15 In practice, a PCA involves 3 steps:

- Choice of pre-processing method, i.e. data pre-treatment method
- Running the PCA algorithm, selection of number of PCs, evaluation of the model
- Interpretation of plots (loading and score plots)

20

Pre-processing is used to ensure that the raw data have a distribution that is optimal for the analysis. Background effects, measurements with various units, different variance of the variables, etc. makes it difficult to extract the meaningful information. Pre-processing reduces the noise introduced by such effects.

25

Pre-processing may contain the operations of centering, weighting and transformations. In the examples transformations have not been used. Transformations include logarithming, smoothing, deriving, normalisation, scatter correction.

30

In (mean-) centering the average of the data of each variable is subtracted. This ensures that all results will be interpreted in terms of variation around the mean. This centering method has been performed in all the analyses/models in the examples.

35

Since PCA is based on finding directions with the largest variation, the result depends on the relative variance of the variables. Depending on the type of information extract from a data set it could be necessary to use a weight based on the variables' standard deviation. However, such weighting is not used in the examples, as the material in question is spectroscopic data, i.e. data of the same origin and the same size.

### Interpretation

Score or loading plots is used in the interpretation of principal component analyses. In these plots the original values are depicted as a function of the principal components. The score plot can be seen as a "map" of the samples whilst the loading plot is, correspondingly, a "map" of the variables.

A plot of PC 1 and PC 2 scores is shown in figure 2.

The score plot shows the clustering of different samples. Samples that lie close to each other in the score plot (sample2 and sample6) are similar with respect to the measured variables. On the other hand, the samples that lie diagonally opposite each other in the plot will be very different from each other (sample3 vs. sample5 or sample2 and sample6 vs. sample1). Additionally, the samples, which lie close to zero in the plot (sample4), are near to the average with regard to the variables measured.

A plot of the first two principal component loadings is shown in figure 3.

The loading plot (Figure 3) is a visualisation of the measured variables that can be attributed the greatest weight when separating the samples. The variables that lie far away from the axes, either in a positive or a negative direction (PARAM.2, PARAM.3, PARAM.4, PARAM.5, PARAM.6 and PARAM.7) will be of greatest importance when separating the samples. Each of the axes in the plot represent a principal component which means that they describe a certain part of the differences in the set of figures, and this is usually depicted in the corner of the plot. The variables that lie far out on the axes with high percentages are most important. That

is, these variables are most prominent in explaining the differences between samples.

5 On the other hand, it also means that variables that lie very close to zero in the plot (PARAM.1) are of no significance to the separation of the samples, i.e. these could be spared. It is thus possible to identify the variables that are dominating to the samples that have been tested.

10 The mutual location of the variables in the loading plot, reveals which variables provide the same information about the samples. Variables that lie on the same straight line through zero in the plot hang together: If they lie on the same side of zero (PARAM.5 and PARAM.6) then there is a positive correlation, i.e. when one is high then the other is too, but if they lie on both sides of zero (PARAM.5 and PARAM.7), then there is a negative correlation, i.e. one is high and the other is low.  
15 Variables that lie like this supplement only each other. They do not contribute additionally to the information concerning the differences between the samples. If, for example, three variables lie on the same spot in the plot, one can be content with the one and still obtain the same information.

20 By comparing the scores and loading plots it can finally be determined which of the variables that determine the grouping of different samples or sample clusters. When the samples lie on the "same spot" in the score plot (relatively, for example in the right corner of the plot) as one or several variables in the corresponding loading plot, the samples have relatively high values for these variables. One can, in other words,  
25 ascertain which variable is the dominating one to the different samples. In the given example the variables, PARAM.5 and PARAM.6, are essential to sample2 and sample6. Likewise, the variable, PARAM.4, is essential to sample3

### 30 Outliers

The definition of an outlier is an observation (outlying sample) or variable (outlying variable) which is abnormal compared to the major part of the data.

35 Extreme points are not necessarily outliers; outliers are points that apparently do not belong to the same population as the others, or that are badly described by a model.

Outliers should be investigated before they are removed from a model, as an apparently outlier may be due to an error in the data.

5 For more information of PCA and multivariate data analysis, see Esbensen KH (2000).

Other types of multivariate analysis that can be used according to the present invention comprises general multivariate analysis, principal component analysis and extensions of this, exploratory and confirmatory factor analysis in its various forms,  
10 Cluster and latent class analysis including scaled latent class analysis, structural equation analysis, Fixed mixture analysis and combinations hereof.

As an alternative to multivariate analysis, the spectral data may be used for training a neural network. Once the neural network has been trained with a sufficiently high  
15 number of known spectra, the neural network is capable of grouping a spectrum from an unknown sample to the correct group, e.g. wildtype; mutant, non-interacting sequence, homozygote heterozygote.

All the methods are related to general mixture analysis (McLouchlan). In PCA, all  
20 the variance is included as total variance, while in factor analysis and most other methods, the variance is portioned into the variance caused by the factor(s) (communality) and unique variance for the indicator. Often the factor(s) nature is not known i.e. is latent. In that sense, factor-analysis is related to latent class analysis (J.A. Hagenaars and A.L. McCutcheon: Applied latent class analysis. Cambridge  
25 University Press, 2002), latent trait analysis (T. Heinen: Latent Class and Discrete Latent Trait Models. Similarities and Differences. Sage Publications, 1996), and the various implementation of these techniques to categorical, ordered and continuous variables. In this context both observed and un-observed and exogenous and endogenous data are referred to. While these techniques are focused on clustering  
30 data in mutual exclusive groups general under the assumption of local independence between indicators, an assumption which can be modified, these analytical tools are presently limited to rather simple structures or models for the interrelationship between data. Structural equation modelling (SEM), in contrast, is capable of handling very complex models with in principal unlimited levels of latent  
35 variable. SEM is however not aimed at clustering data as latent class or latent

cluster analysis, but rather as a tool to develop and verify causal interrelationships between items, events etc. These two groups of techniques are therefore complementary rather competitive. In the present context PCA is used for basic data-analyses and reduction. In further analysis of the spectra, in particular when two or more labels are used, the above-mentioned techniques will be applicable in data-treatment and presentation. E.g. if two labels are used, spectre from two labels will be obtained, which preferable should be analysed separately. Clustering of the spectre, analysis of interaction between the spectre of each of the labels concomitantly obtained and presenting data in a comprehensible and interpretable way will use elements of all the above mentioned techniques.

### **Spectroscopy**

Spectroscopic techniques form the largest and most important single group of techniques used in analytical chemistry, and provide a wide ranges of quantitative and qualitative information. All spectroscopic techniques depend on the emission or absorption of electromagnetic radiation characteristic of certain energy changes within an atomic or molecular system. The energy changes are associated with a complex series of discrete or quantised energy levels in which atoms and molecules are considered to exist.

The relevant spectroscopic techniques included but are not limited to:

#### **Table 1**

25	arc/spark spectrometry
	spectrography
	plasma emission spectrometry
	flame photometry
	X-ray fluorescence spectrometry
30	atomic fluorescence spectrometry
	atomic absorption spectrometry
	$\gamma$ -spectrometry
	ultraviolet spectrometry
	visible spectrometry
35	infrared spectrometry

nuclear magnetic resonance spectrometry  
mass spectrometry

5 Preferably the spectroscopy comprises ultraviolet spectrometry, visible spectrometry or infrared spectrometry. These spectroscopic methods can all be used in conjunction with numerous known labels and the equipment used for recording the spectra is standard laboratory equipment and thus generally available.

10 Recording a spectrum under normal conditions comprises recording a value (absorption, extinction, emission etc) for a number of discrete wavelengths, because spectrometers normally do not record continuous spectra. In order to be able to have as much data as possible for the multivariate analysis, the recording of a spectrum comprises recording at as many discrete wavelengths as possible or recording a continuous spectrum if possible. Thus recording spectral data preferably  
15 comprises detection of signal for at least 10 discrete wavelengths, more preferably at least 20 discrete wavelengths, more preferably at least 50 discrete wavelengths, more preferably at least 100 discrete wavelengths, more preferably at least 200 discrete wavelengths, more preferably at least 250 discrete wavelengths, more preferably at least 300 discrete wavelengths, more preferably at least 400 discrete  
20 wavelengths, more preferably at least 500 discrete wavelengths, more preferably at least 600 discrete wavelengths, more preferably at least 750 discrete wavelengths, more preferably at least 1000 discrete wavelengths, such as at least 1250 discrete wavelengths, for example at least 1500 discrete wavelengths, such as at least 2000 discrete wavelengths.

25 In order to be able to have as high a resolution as possible in the spectrum, the distance between the different wavelengths preferably is as low as possible. This increases the possibility of resolving the spectra into different groups. Accordingly, the distance between the discrete wavelengths preferably is 10 nm or less, more  
30 preferably 5 nm or less, more preferably 3 nm or less, more preferably 2 nm or less, more preferably 1 nm or less, more preferably 0.8 nm or less, more preferably 0.75 nm or less, more preferably 0.7 nm or less, more preferably 0.6 nm or less, more preferably 0.5 nm or less, more preferably 0.25 nm or less, more preferably 0.1 nm or less, more preferably 0.05 nm or less, more preferably 0.01 nm or less.

35

In further preferred embodiment of the method of the invention, the spectral data recorded comprises a fluorescence spectrum between 180 and 950 nm.

5 A fluorescence spectrum may be an excitation spectrum or an emission spectrum, or both, depending on the level of reduction of variance not related to the interaction of target and non-target. To the extent that these non-interactive variances can be eliminated e.g. by clustering the spectre both emission and excitation spectre alone or in combination can be of analytical value.

10 In order to better analyse the difference between a spectrum from an unbound and a bound probe, the method preferably further comprises recording of spectral data from the polynucleotide probe alone. After having established this difference, it is expected that it is not always necessary to record this spectrum, although it may in some instances be useful as a calibration for day-to-day variation.

15 The method may further comprise recording spectral data from the hybrid polynucleotide and from a polynucleotide probe alone and/or, from a non-hybridising polynucleotide probe contacted by the target polynucleotide, and/or from a polynucleotide probe contacted with a non-hybridising polynucleotide sequence. All  
20 of these spectra, which may be regarded as "controls" serve the purpose of calibrating the method.

Mass spectrometry is a technique for characterising molecules according to the manner in which they fragment when bombarded with high-energy electrons, and for  
25 elemental analysis at trace levels. It is not strictly speaking a spectrometric method as electromagnetic radiation is neither absorbed nor emitted. However, the data obtained are in a spectral form in that the relative abundance of mass fragments from a sample is recorded as a series of lines or peaks. The bombardment process produces many fragments carrying a charge, and this facilitates their separation and  
30 detection by electrical and magnetic means. Spectra must be recorded under conditions of high vacuum ( $10^{-4}$  to  $10^{-6}$  Nm<sup>-2</sup>) to prevent loss of the charged fragments by collision with molecules of atmospheric gases or swamping of the sample spectrum.



In the case of using mass spectrometry for recording the spectrum, it is contemplated that it may not always be necessary to include a label in the polynucleotide probe, since the association between probe and target alone is enough to create differences in the recorded spectrum.

5

## Labels

10

In the present context, the term "label" means a chemical moiety which is coupled to a nucleic acid of a polynucleotide probe (possibly via a molecular linker) and which can be used as a signal source for electromagnetic radiation or as a source of interaction with electromagnetic radiation supplied to the label. The polynucleotide probes may be labelled by a number of methods well known in the art. Conveniently, polynucleotide probes may be labelled during their solid-phase synthesis using any of the many commercially available phosphoramidite reagents for 5' labelling. Illustrative examples of oligonucleotide labelling procedures may be found in US 6,255,476.

15

20

A preferred label according to the invention has a fairly complex spectrum, which when resolved at short wavelength distances produces more than one local maximum in addition to the global maximum. When interacting with one target polynucleotide the recorded spectrum preferably changes at more than one wavelength. Thereby more data can be accumulated and used for the multivariate analysis.

25

A fluorescent, phosphorescent or luminescent label is preferred because it provides a strong signal with low background. It is also optically detectable at high resolution and sensitivity through a quick scanning procedure. More preferably the label comprises a fluorescent label.

30

A particular fluorescent label has a characteristic excitation and emission spectrum which allows the simultaneous detection of several different fluorescent labelled molecules if the labels are selected appropriately.

35

A large number of different useful fluorescent labels are given in the art and may be selected from the group comprising, but not limited to: Cy2, Cy3, Cy3.5, Cy5, Cy5.5,

Cy7 (trademarks for Biological Detection Systems, Inc.), fluorescein, acridin, acridin orange, Hoechst 33258, Rhodamine, furthermore: Rhodamine Green, Tetramethylrhodamine, Texas Red, Cascade Blue, Oregon Green, Alexa Fluor (trademarks for Molecular Probes, Inc.), 7-nitrobenzo-2-oxa-1-diazole (NBD), pyrene and Europium, Ruthenium, Samarium, and other rare earth metals, bimeane, ethidium, europium (III) citrate, La Jolla blue, methylcoumarin, nitrobenzofuran, pyrene butyrate, rhodamine, and terbium chelate. More specialised fluorochromes are listed in Table 2 along with their suppliers.

10

TABLE 2  
FLUORESCENT LABELS

	Fluorochrome	Vendor	Absorption Maximum	Emission Maximum
	Bodipy 493/503	Molecular Probes	493	503
15	Cy2	BDS	489	505
	Bodipy FL	Molecular Probes	508	516
	FTC	Molecular Probes	494	518
	FluorX	BDS	494	520
	FAM	Perkin-Elmer	495	535
20	Carboxyrhodamine	Molecular Probes	519	543
	EITC	Molecular Probes	522	543
	Bodipy 530/550	Molecular Probes	530	550
	JOE	Perkin-Elmer	525	557
	HEX	Perkin-Elmer	529	560
25	Bodipy 542/563	Molecular Probes	542	563
	Cy3	BDS	552	565
	TRITC	Molecular Probes	547	572
	LRB	Molecular Probes	556	576
	Bodipy LMR	Molecular Probes	545	577
30	Tamra	Perkin-Elmer	552	580
	Bodipy 576/589	Molecular Probes	576	589
	Bodipy 581/591	Molecular Probes	581	591
	Cy3.5	BDS	581	596
	XRITC	Molecular Probes	570	596
35	ROX	Perkin-Elmer	550	610
	Texas Red	Molecular Probes	589	615

	Bodipy TR (618)	Molecular Probes	596	625
	Cy5	BDS	650	667
	Cy5.5	BDS	678	703
	DdCy5	Beckman	680	710
5	Cy7	BDS	443	767
	DbCy7	Beckman	790	820

The suppliers listed in Table 2 are Molecular Probes (Eugene, Oreg.), Biological Detection Systems ("BDS") (Pittsburgh, Pa.) and Perkin-Elmer (Norwalk, Conn.).  
 10 Several forms and colours of CyDye fluorophores are currently available; the spectral properties for these dyes are listed in Table 3.

Table 3: The spectral properties of CyDye Fluors							
Fluorophore	Colour of Fluorescence	Absorption max (nm)	Fluorescence max (nm)	Extinction Coefficient (M-1cm-1)	Quantum yield for protein conjugates	Formula Weight (Daltons)	
						Bisfunctional	Mono-functional
Cy2	Green	489	506	~150,000	>0.12	896.95	713.78
Cy3	Orange	550	570	150,000	>0.15	949.11	765.95
Cy3.5	Scarlet	581	596	150,000	>0.15	1285.54	1102.37
Cy5	Far-red	649	670	250,000	>0.28	975.15	791.99
Cy5.5	Near IR	675	694	250,000	>0.28	1311.58	1128.41
Cy7	Near IR	743	767	~250,000	~0.28	1001.19	818.02
FluorX	Green	494	520	68,000	0.3		586.6

In is at present preferred to use a Cy dye as detectable labels according to the  
 15 present invention, more preferably the label comprises a Cy5 dye.

### Quenchers

In some embodiments of the present invention a quencher is added to the sample  
 20 with the hybrid between target and probe. This may be preferable in certain cases in order to quench parts of the spectrum, which only contributes to noise. This would

be most appropriate when non-relevant variance (noise) cannot be eliminated by multivariate analysis. In some instances quenchers may act as modifiers to alter the general spectral characteristics.

- 5 Examples of quenchers depend on the specific part(s) of the spectrum to be quenched. The quenching abilities of a number of different compounds are known in the art. One particular example is the TAMRA quencher.

#### **Immobilisation to solid surface**

10

In some embodiments of the present invention it is preferable that one or more probes or targets polynucleotide is immobilised to a solid surface. The nature of the means for immobilisation and of the nature of the solid support is a matter of choice. Numerous suitable supports and methods of attaching nucleotides to them are well known in the art and widely described in the literature. Thus for example, supports in the form of microtiter wells, tubes, dipsticks, particles, fibers or capillaries may be used, made for example from agarose, cellulose, alginate, teflon, latex or polystyrene. Conveniently, the support may comprise magnetic particles, which permits the ready separation of immobilised material by magnetic aggregation.

20

The solid support may carry functional groups such as hydroxyl, carboxyl, aldehyde or amino groups for the attachment of nucleotides. These may in general be provided by treating the support to provide a surface coating of a polymer carrying one of such functional groups, e.g. polyurethane together with a polyglycol to provide hydroxyl groups, or a cellulose derivative to provide hydroxyl groups, a polymer or copolymer of acrylic acid or methacrylic acid to provide carboxyl groups or an amino alkylated polymer to provide amino groups. US 4,654,267 describes the introduction of many such surface coatings.

25

- 30 Alternatively, the support may carry one member of an "affinity pair", such as avidin, while the polynucleotide is conjugated to the other member of the affinity pair *in casu* biotin. Representative specific binding affinity pairs are shown in Table 4.

35

TABLE 4. Representative Specific Binding Pairs	
antigen	Antibody
hapten (e.g. digoxigenin)	anti-hapten antibody or Fab-fragment
biotin	avidin (or streptavidin or anti-biotin)
IgG*	protein A or protein G
5-Bromo-2'-Deoxyuridine	specific antibodies
drug	drug receptor
toxin	toxin receptor
carbohydrate	lectin or carbohydrate receptor
peptide	peptide receptor
protein	protein receptor
enzyme substrate	enzyme
DNA (RNA)	aDNA (aRNA)**
hormone	hormone receptor
ion	chelator
<i>lac</i> repressor protein Lac I	<i>lac</i> operator ( <i>LacOP</i> )
*IgG is an immunoglobulin.	
**aDNA and aRNA are the antisense (complementary) strands used for hybridisation.	

5 The streptavidin/biotin binding system is very commonly used in molecular biology, due to the relative ease with which biotin can be incorporated within nucleotide sequences, and indeed the commercial availability of biotin-labelled nucleotides, and thus biotin represents a preferred means for immobilisation.

10 In a preferred embodiment of the invention, an amplified DNA strand is labelled with a molecule, which is subsequently used to immobilise the labelled DNA strand to a solid surface.

15 The target polynucleotide may be labelled by a number of methods. One convenient method to label DNA is to enclose one labelled amplification primer oligonucleotide in the amplification reaction mixture. During the amplification process the labelled oligonucleotide is built into the DNA fragment which become labelled.

During synthesis oligonucleotides may also be labelled or coupled to chemoreactive groups comprising, but not limited to: sulfhydryl, primary amine or phosphate. The subsequent use of such labelled primers in PCR, LCR or similar oligonucleotide dependent amplification methods results in labelled DNA fragments which can be immobilised on specialised surfaces. For instance SH-modified DNA may be immobilised on a gold surface (Steel et al. (2000) Biophys J 79:975-81) likewise 5'-phosphorylated DNA or 5'-aminated DNA may be immobilised by reaction with activated surfaces (Oroskar et al. (1996) Clin Chem 42:1547-55; Sjoroos et al (2001) Clin Chem 47:498-504).

During synthesis oligonucleotides may also be labelled or coupled to photoreactive groups. Acetophenone, benzophenone, anthraquinone, anthrone or anthrone-like modified DNA can for instance be activated by exposure to UV light and immobilised on a wide range of surfaces as described in European and US patents: EP0820483, US6033784 and US5858653. Also photoreactive psoralens, coumarins, benzofurans and indols have been used for immobilisation of nucleic acids. An extensive discussion of immobilisation of nucleic acids can be found in WO 85/04674.

## **Target polynucleotides**

A target polynucleotide is defined as a polynucleotide with which at least one probe can form a hybrid polynucleotide.

The methods, kits and systems of the present invention can be used with any kind of polynucleotide, which can be subjected to a hybridisation assay. One source of target polynucleotide are RNA polynucleotides, such as wherein the target polynucleotide comprises RNA, such as mRNA and/or rRNA and or tRNA.

Among the RNAs the rRNAs including 5S, 5.5-5.8S, 16S, 18S, 23S, 25-28S rRNA can be used for identifying the taxon from which the RNA was isolated, since conserved sequences can be found for a number of taxons and for a number of species. rRNA may for example be used for diagnosing an infectious disease caused by microorganisms, or for determining the amount and nature of contamination in food, feed and various water-supplies.

Another source of target polynucleotides are DNA. DNA may be used for determination of mutations and/or polymorphisms, but also in genotyping individuals or for determining the taxon, for forensic usage or for linkage studies. The different  
5 sources of DNA include genomic DNA, organelle DNA, mitochondrial DNA, chloroplast DNA, cDNA, and environmental DNA.

The target polynucleotides may also comprises a synthetic polynucleotide sequence. For example it may be advantageous to include a restriction site in the  
10 target polynucleotide so that the sequences can be cloned into a vector afterwards. Mutations may also be included in the sequences during isolation and/or amplification, e.g. via so-called error-prone PCR.

Further examples of synthetic sequences include but are not limited to the inclusion  
15 of various control polynucleotides in the hybridisation mixture, such as positive controls (wild-type, mutation, heterozygote), negative control (dummy DNA sequence).

In certain cases the target polynucleotide comprises chemically or biologically  
20 modified nucleic acids, such as the modification comprising modification of cytosin by bisulphite. An example of this is the study of methylation patterns usually in promoters and enhancers.

The target polynucleotide may comprise a mixed polymer of any of the polymers  
25 described above. For example the target polynucleotide may comprise a cyclic RNA/DNA polymer of the type used for gene therapy.

The length of the target polynucleotide is not important for the function of the  
invention. The target polynucleotide may thus comprise from 8 to 50,000 bases or  
30 even more than 50,000 bases. The choice of length is mainly dependent on the purification and amplification steps, the number of sub-sequences to interrogate with the probes and the distance between such sub-sequences, which may each comprise a polymorphism. However, usually the target need not exceed the length of the probes with more than 5 nucleotides in the terminal labelled end.

35

- The length of the target polynucleotide may thus be selected from 8-15 bases, from 15-30 bases, from 30 to 50 bases, from 50 to 100 bases, from 100 to 200 bases, from 200 to 300 bases, from 300 to 500 bases, from 500 to 750 bases, from 750 to 1000 bases, from 1000 to 1500 bases, from 1500 to 3000 bases, from 3000 to 5000 bases, from 5000 to 10000 bases, from 10000 to 15000 bases, from 15000 to 20000 bases, from 20000 to 25000, from 25000 to 30000 bases, from 30000 to 35000 bases, from 35000 to 40000 bases, from 40000 to 45000 bases, from 45000 to 50000 bases, more than 50000 bases.
- When referring to the length of the target polynucleotide, the whole length of the molecule is intended. During hybridisation, the probe only hybridises to one or more relatively short sequence(s) of the target polynucleotides. These sub-sequences to which the probe hybridises are termed the range of overlap between target and probe. This length of the overlap between the probe and target polynucleotide may be as short as at least 5 nucleotides. However more specific hybridisation is obtained by increasing the length of the overlap. Therefor, more preferably the overlap is at least 6 nucleotides, such as at least 7 nucleotides, for example 8 nucleotides, such as at least 9 nucleotides, for example at least 10 nucleotides, such as at least 15 nucleotides, for example at least 20 nucleotides, for example at least 25 nucleotides, such as at least 50 nucleotides, for example at least 100 nucleotides. In other preferred embodiments, the overlap is at most 100 nucleotides.

#### **Extraction of target nucleic acids**

- Extraction of target nucleic acids can be performed using methods known to those skilled in the art (Joseph Sambrook & David W Russell (2001) "Molecular cloning: a laboratory manual", Cold Spring Harbor Laboratory Press, New York, USA). When selecting extraction protocols among the numerous available protocols, it is preferable to select protocols where the buffer ingredients do not interfere with the recordation of spectral data. However, it is also possible to precipitate the extracted target nucleic acid to get rid of components, which interfere with the spectrum.

#### **Probes**



The probes used in the method according to the present invention may be made from any kind of nucleotide monomer or any combination of the known types of monomers. Thus the probe may comprise at least one RNA monomer, and/or comprises at least one DNA monomer, and/or at least one PNA monomer, and/or at least one methylated monomer, and/or at least one labelled monomer, and/or at least one LNA monomer.

The probe is designed to hybridise specifically to a predetermined sequence in the target polynucleotide if this sequence is present there. Therefore, the probe may be designed to hybridise to any sequence of interest.

One area of particular interest for the present invention is the detection of polymorphisms and/or mutations related to human diseases or animal diseases. Any disease or health related problem influenced by genetic factors is envisaged, such as any one or more listed in the International Statistical Classification of Disease and Health Related Problems, ICD-10, of the World Health Organisation. In the following a number of such human diseases with a genetic impact are described. In the sequences disclosed below, the polymorphism is located at the centre of the sequences. In most of the disclosed sequences, the nucleotide complementary to the polymorphism is located approximately at the centre of the probe as is usually done with probes for detection of polymorphisms. The label is most often bound to either the 3' or 5' terminal nucleotide. The label may also be bound to another nucleotide, e.g. it may be bound to a (non-terminal) nucleotide being complementary to the polymorphism.

When designing probes for use according to the present invention, it is more advantageous to have the nucleotide complementary to the polymorphism positioned terminally than centrally in the probe. When the polymorphism is located terminally, the end of the probe will either be completely complementary to the target nucleic acid sequence and hybridise at all positions, or the nucleotide in the end will be non-complementary and therefore probably bend away from the target nucleic acid sequence. When the label is also bound to this terminal nucleic acid, which is either complementary or non-complementary to the target, then the spectral difference in signal between the hybrid with the wild-type and the hybrid with the mutant target nucleic acid sequence is maximised.

Another possibility is to design a probe, which does not overlap with the polymorphic site, and which has the label in the position complementary to the polymorphic site. Such a probe will give rise to a spectral difference in signal between wild-type and mutant, since the label will interfere differently with the different nucleic acids in the site.

A further possibility is to have the label bound to a non-terminal nucleotide, which is complementary to the polymorphic site. When using such probes, the spectral difference between wild-type and mutant is also enhanced.

Generally, it can be said that those probes are preferred, where the label is positioned as close as possible to the polymorphic site to maximise the spectral difference. The distance from the label to the polymorphic site may be 1 nucleotide, 2 nucleotides, 3 nucleotides, 5 nucleotides, 10 nucleotides or more. Among these, the most preferred are those where the nucleotide complementary to the polymorphic site is in a terminal position.

apolipoprotein B mutations related to atherosclerosis, wherein the probe may comprise a sequence from any of SEQ ID NO 1 to 4.

apolipoprotein E polymorphism (apoE2, E3 and E4) related to neurological diseases, wherein the probe may comprise a sequence from any of SEQ ID NO 5 to 8.

human muscle glycogen synthase polymorphism related to diabetes mellitus, wherein the probe may comprise a sequence from any of SEQ ID NO 9 to 10.

methylene tetrahydrofolate reductase polymorphism related to osteoporose, wherein the probe may comprise a sequence from any of SEQ ID NO 13 to 14.

Dnase1 mutations related to rheumatological diseases, wherein the probe may comprise a sequence from any of SEQ ID NO 11 to 12.

BRCA1 gene or in the BRCA2 gene related to breast or ovarian cancer, wherein the probe may comprise a sequence from any of SEQ ID NO 27 to 30.

5 Mismatch repair gene mutations related to cancer, wherein the probe comprises a sequence from any of SEQ ID NO 15 to 16.

The probe may be selective for a mutation in a promoter sequence, or the probe may be selective for a mutation in a coding sequence, including introns and exons.

10 Apart from mutations, the probes may be used to diagnose the presence/absence and/or nature of a microbial infection, wherein the probe is selective for a microbial target nucleic acid sequence. Preferably the probe is selective for a microbial 16S, 18S, or 23S rRNA sequence, because these contain sequences which are conserved across a large group of microbes as well as sequences, which are  
15 species-specific. SEQ ID NO 17 is a general probe, which captures 16S RNA. SEQ ID NO 18 is an example of a probe which is specific for all Enterobacteriaceae 16S. SEQ ID NO 19 is a species specific probe which detects E. coli ECA75F. As can be understood from the foregoing, it is possible for the skilled person to design the necessary probes for any desired purpose.

20

#### **Layout of assays**

It is contemplated that the method of the present invention can be used for multiple hybridisation assays in the same vessel, because the contribution to the total  
25 spectrum by the various hybrids formed can be resolved by the multivariate analysis. Accordingly, one can use at least two polynucleotide probes capable of hybridising to two different target polynucleotides. Preferably, but not necessarily the two probes are linked to two different detectable labels.

30 It is also contemplated that one may use at least three probes capable of hybridising to three different target polynucleotides or even four or five or more probes capable of hybridising to four, five or more different target polynucleotides. Preferably, the three, four, five or more probes are linked to three, four, five or more different detectable labels.

35

One special embodiment of the multi-probe layout comprises the use of at least one probe having at least two stretches of complementarity to at least one target polynucleotide, such as at least 3 stretches, for example at least 4 stretches, such as at least 5 stretches. Preferably, the two stretches are separated by a nucleotide sequences, which does not hybridise to the target polynucleotide. Such probes are adapted e.g. for determination of the haplotype of multiple mutations/polymorphisms in the same gene.

In the simplest version of the method, hybridisation and recording of spectral data is performed in solution. This is because it is not necessary to purify the hybridised polynucleotides and remove e.g. unbound probe. Therefore it is possible to record spectral data from a solution comprising both the hybrid polynucleotide and unhybridised probe. The spectrum from the complex solution is changed by the mere presence of the hybrid.

The ratio of target to probe in the hybridisation solution may range from 1:0.1 to 1:10, such as 1:0.2, 1:0.5, 1:0.75, 1:1; 1:2, 1:4, 1:5, 1:7, 1:8, or 1:10.

However, it is likewise possible to record spectral data from hybrid polynucleotides bound to a solid support. The solid support may comprises a solid surface capable of immobilising a capture probe, a capture probe capable of immobilising the target polynucleotide, and a labelled detection probe capable of hybridising to the immobilised target polynucleotide. The capture probe may be is immobilised a priori to the solid surface or the capture probe may be hybridised to the target before immobilisation on a solid support. In a separate embodiment the capture probe(s) is/are (an) allele specific probe(s).

In a special layout the solid support is a disposable or reusable device such as but not exclusively a flow-through system. Such a flow through system may comprise immobilised capture probes to capture the target, which can then be labelled by hybridisation with a label probe. However, it is also possible to label the capture probe directly, and obviate the need for a separate label probe.

#### Amplification

The target polynucleotides may be amplified by one of many methods. One of the best known and widely used amplification methods is the polymerase chain reaction (referred to as PCR) which is described in detail in US 4,683,195, US 4,683,202 and US 4,800,159, however other methods such as LCR (Ligase Chain Reaction, see  
5 Genomics (1989) 4:560-569), NASBA (Nucleic Acid Sequence-Based Amplification, see *PCR Methods Appl* (1995) 4, S177-S184), strand displacement amplification (*Current Opinion in Biotechnology* (2001) 12:21-27) or rolling circle amplification (*Current Opinion in Biotechnology* (2001) 12:21-27) can be applied. Genuine amplification in e.g. bacteria or yeast can also be used to amplify nucleotide  
10 sequences as can non-PCR methods, such as T7-polymerase.

In the case of polymorphisms being situated a long distance from each other and in certain other cases fairly long fragments of genomic DNA or other polynucleotides have to be amplified. Several techniques that results in amplification of fairly long  
15 fragments of DNA is described in the art. One particular useful procedure is the so-called "long range PCR" which allow amplification of very large fragments (*Proc. Natl. Acad. Sci. USA*. (1994) 91: 2216-20; *Methods-Mol-Biol.* (1997) 67: 17-29.). Kits allowing the amplification of templates up to 40.000 base pair long are commercial available, e.g. the TripleMaster™ PCR System (cat. no. 0032 008.208) of Eppendorf  
20 AG, Hamburg, Germany.

### Hybridisation

The polynucleotide probe is then allowed to hybridise to the target polynucleotide. In  
25 the scope of the present invention the term "hybridisation" signifies hybridisation under conventional hybridising conditions, preferably under stringent conditions, as described for example in Sambrook et al., *Molecular Cloning, A Laboratory Manual*, 2nd Edition (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.). The term "stringent" when used in conjunction with hybridisation conditions is as  
30 defined in the art, i.e. 15-20°C under the melting point  $T_m$ , cf. Sambrook et al, 1989, pages 11.45-11.49. Preferably, the conditions are "highly stringent", i.e. 5-10°C under the melting point  $T_m$ . Under highly stringent conditions hybridisation only occurs if the complementarity between a sequence of the polynucleotide probe and a sequence of the target polynucleotide of interest is 100 %, while no hybridisation  
35 occur if there is just one mismatch. Such optimised hybridisation results are reached

by adjusting the temperature and/or the ionic strength of the hybridisation buffer as described in the art. However equally high specificity may be obtained using high-affinity DNA analogues. One such high-affinity DNA analogues has been termed "locked nucleic acid" (LNA). LNA is a novel class of bicyclic nucleic acid analogues in which the furanose ring conformation is restricted in by a methylene linker that connects the 2'-O position to the 4'-C position. Common to all of these LNA variants is an affinity toward complementary nucleic acids, which is by far the highest reported for a DNA analogue (Ørum et al. (1999) *Clinical Chemistry* 45, 1898-1905; WO 99/14226 Exiqon). Another high-affinity DNA analogues is the so-called protein nucleic acid (PNA). In PNA compounds, the sugar backbone of an oligonucleotide is replaced with an amide containing backbone, in particular an aminoethylglycine backbone. The nucleobases are retained and are bound directly or indirectly to aza nitrogen atoms of the amide portion of the backbone (*Science* (1991) 254: 1497-1500).

In general, the hybridisation conditions are chosen so that the formation of a hybrid polynucleotide takes place under conditions of optimal or suboptimal stringency providing sufficient stable complexes for discriminatory signal detection,

any composition of buffers optimising discriminatory signal detection,

any form and concentrations of one or more salts optimising discriminatory signal detection,

any additives including but not limited to stabilisers and/or quenchers optimising discriminatory signal detection,

temperature range for hybridisation specific for any specific combination of analyte and probe optimising discriminatory signal detection,

any range of time of hybridisation necessary to optimise discriminatory signal detection.

As evidenced by the appended examples the hybridisation temperature can be varied within the normal limits. In other words, the formation of a hybrid may be performed at a temperature between 10 and 90°C such as 10 to 20 °C, 20-30 °C, 30 to 40 °C, 40 to 50 °C, 50 to 60 °C, 60 to 70 °C, 70 to 80°C, or 80 to 90°C.

The buffer used for the hybridisation is conveniently a PCR buffer, which preferably is non-fluorescent, and/or which stabilises the spectrum of electromagnetic radiation, and which allows hybridisation.

## 5      **Determination of haplotypes**

10      The methods of the present invention may likewise be used for the determination of haplotypes. The haplotype is the set of alleles borne on one of a pair of homologous chromosomes. Often the particular combination of alleles in a defined region of some chromosome, e.g. the locus of the major histocompatibility complex (MHC), is referred to as the haplotype of that locus. The central dogma of modern molecular genetics teaches that it is the haplotype of the coding part of a gene that determines the amino acid sequence and thus the function of the resulting protein. In connection with studies aimed at establishing an association between the risk of developing a  
15      particular disease and the genetic makeup of the patients, information of the haplotype-investigations are superior to conventional genetic investigations of only single loci, because the haplotype provides physiological relevant information from more loci.

20      The determination of the haplotype may for example be carried out using allele specific primers to amplify only one of two alleles. This is done by using a primer set, wherein at least one of the primers is 100% complementary to only one allele and carrying out the annealing step of the amplification at high stringency, so that hybridisation will only take place between the primer and the allele being 100 %  
25      complementary to the primer. The presence of a further polymorphism in the amplified fragment can be determined using an allele specific polynucleotide probe.

30      In another embodiment both alleles are isolated and amplified (if required) and the two strands are separated using allele specific capture probes. The presence of a further polymorphism on the separated fragments can be determined using an allele specific polynucleotide probe.

35      In further embodiments of the invention both the resolution of the haplotypes and the detection of the multiplicity of nucleotide polymorphisms are performed in suspension, defined here as a "one-phase system".

In one embodiment this is accomplished by the amplification of the preselected region. The amplified fragment is hybridised to specially designed probes which are capable of detecting a multiplicity of polymorphisms. Such "bifunctional" probes contain at least 2 stretches of relatively short sequences which are complementary to each of the two studied polymorphisms separated by a region of spacer DNA which will not hybridise with the amplified region under the conditions in the experiment.

To obtain hybridisation signal from only one chromosome of a chromosome pair, the procedure takes advantage of the fact that intramolecular hybridisation is thermodynamically favorable compared to hybridisation between separate molecules. This difference between intra- and inter-molecular hybridisation is in particular significant when hybridisation is performed at low concentrations of nucleic acid and results in a significant difference even between the hybridisation of the bifunctional probe to one or two amplified fragments, the hybridisation to one fragment being the most favourable. Under stringent conditions employed in the hybridisation only probes which are completely complementary to the sequences comprising both studied polymorphisms will form stable hybrids.

Following the stringent hybridisation, the haplotype can be determined by recording a fluorescence spectrum from the different oligonucleotides hybridised to the amplified DNA without the need for separation of hybrid from .

Further details on how to perform determination of haplotypes can be found in PCT/DK02/00552 (HVIDOVRE HOSPITAL), which is hereby incorporated by reference in its entirety.

#### **Kits for detection of mutations/polymorphisms**

The invention also relates to a kit for detection of mutations or polymorphism comprising

at least one oligonucleotide probe capable of hybridising to a preselected region of a target polynucleotide, the polynucleotide probe further comprising at least one detectable label,



instructions enabling correlation of spectral data recorded from a hybrid polynucleotide between said at least one oligonucleotide probe and said target polynucleotide to the presence or absence of said mutation or polymorphism using multivariate analysis.

5

The kit is assembled and sold with probes and instructions enabling the performance of one of a number of different assays as outlined above. The instructions can be regarded as calibration data, which enable the user to perform the hybridisation assay without first determining the difference between a spectrum from an unbound and a bound probe. The instructions are generally rather complex and may conveniently be stored on a data storage medium, which is provided together with the probes, and optional buffers. Instructions may be in the form of calibration data on a data carrier, such as floppy disc, a CD-ROM, a DVD, ROM, chips, memory-cards, bar-codes. Instructions in the form of calibration data may also be transferred over a network and the kit may then comprise the address of a propagated signal comprising calibration data, which can be transferred over a network, such as e-mail, internet, on-line nets, fibre-optics, power-cables, satellite-dishes.

20 Finally, it may be possible in some cases to provide instructions in the form of calibration data which can be entered into a computer unit.

The oligonucleotide probes comprised in the kit may also in one embodiment be a multifunctional probe having two or more sequences which can hybridise to separate sequences in the target. The two or more sequences are separated by a spacer sequence, which does not hybridise to the target. Examples of such multifunctional probes can be found in PCT/DK02/00552 (Hvidovre Hospital).

25

The kits may also comprise two or more differentially labelled probes, which hybridise to different sequences in the target nucleotide. Such differently labelled probes can be used for determining the linkage phase of two or more mutations/polymorphisms.

30

Often the kits will additionally comprise at least one control polynucleotide capable of hybridising to the oligonucleotide probe and non-hybridising polynucleotide(s).

35

5 The probes in the kits can be in the form of probes, which can be used in solution or the probes may be linked in one end to a moiety, which may be used for immobilising the probes to a solid surface, as described above. The kit may also be in the form of a tube with at least one probe linked to the inner surface, the tube wall allowing electromagnetic radiation to pass the walls. The tube is made in a shape which fits into a standard spectrophotometer.

10 Such a tube may comprises more than one probe linked to more than one location, the locations being spatially separate. For example the tube may comprise a probe selective for a wildtype in one location and a mutation/polymorphism in another location. The tube may also comprise more than one probe the probes having detectably different labels, which preferably do not interfere with each other's spectrum. In this way hybridisation to the two probes can be detected with one scan  
15 over a given wavelength range.

#### **Systems for detecting a hybrid polynucleotide**

20 In a further aspect the invention relates to a system for detection of a hybrid polynucleotide comprising  
at least one oligonucleotide probe being at least partly complementary to a target polynucleotide, the probe comprising a detectable label,  
a sample chamber from which electromagnetic radiation can be recorded,  
a source of spectrally resolved electromagnetic radiation,  
25 means for sensing and recording a spectrum of electromagnetic radiation from the sample chamber,  
a computer unit for storing spectral data of electromagnetic radiation and having instructions to treat the recorded spectral data using multivariate analysis.

30 These systems are adapted for high-throughput screening of hybridisation events.

The system may further comprising a computer controlled robot to transfer solutions to the sample chamber.

Preferably the system comprises means to control the temperature of the sample chamber during hybridisation and subsequent recording of spectrum.

5 As for the kits, the sample chamber may be in the form of a tube with at least one probe linked to the inner surface. The tube may comprise more than one probe linked to more than one spatially separate location and the system comprises means to record a spectrum from each of the spatially separate locations. And the tube may comprise more than one probe, the more than one probe having detectably different labels.

10

The system preferably is made to accommodate standard laboratory glass and/or plasticware, and may in one preferred embodiment be adapted to accommodate a multi-well dish and record a spectrum for each well. The multi-well dish may be a 96 well dish a 384 well or a standard dish with more wells.

15

A solid support may also comprise a dish or a rod, spinning dishes or rotating and displaceable rods.

20

Capture-probes are attached to the dish or the rod in predefined areas so that the hybridised samples are unequivocally identified. Only one or a few spectral detection-units are needed as the spinning of the dish or rod will bring all the samples into focus of the detection device.

## Examples

### Methods

#### 5 Oligonucleotides

The oligonucleotides used for the examples were selected from a region of the Apolipoprotein gene - Apo3611. The oligos are aligned in Figure 4 to show the regions of complementarity.

#### 10 Unlabelled oligonucleotides

Wildtype 1 (SEQ ID NO:20)

**W1**

5'CTAAGAACCAGAAGATCAGATGGAAAAATGAAGTCCGGATTCATTCTGGGTCTTTCCAGA  
GCCAGGTCGA

15

Wildtype 2 (SEQ ID NO: 21)

**W2** 5'ACCAGAAGATCAGATGGAAAAATGAAGTCCGGATTCATTCTGGGTCTTTC

Wildtype 3 (SEQ ID NO: 31)

20

**W3**

5'GCTAACACTAAGAACCAGAAGATCAGATGGAAAAATGAAGTCCGGATTCATTCTGGGTCT  
TTC

Mutant (SEQ ID NO: 22)

25

**M** 5'ACCAGAAGATCAGATGGAAAAATGAAGTCCAGATTCATTCTGGGTCTTTC

Reverse oligonucleotide (SEQ ID NO: 32)

**R** 3'TGGTCTTCTAGTCTACCTTTTACTTCAGGCCCAAGTAAGACCCAGAAAG

30

**Y** Double stranded target molecule formed by W2 and R.

Non-specific target molecule (SEQ ID NO: 33)

**Z** 5'GTTCACGAGCTCAGCAACCTGTGACCTGAATTCAGTCTGATAAAATCGCA

35

#### Cy5 labelled oligonucleotides

**A** (~~Cy5-wt-ter~~)

Cy5-5'-CGGACTTCATTTTTC (SEQ ID NO: 23)

**B** (**Cy5-mu-ter**)

Cy5-5'-TGGACTTCATTTTTC (SEQ ID NO: 24)

**C** (**Cy5-wt-cen**)

Cy5-5'-ATGAATCCGGACTTC (SEQ ID NO: 25)

5 **D** (**Cy5-mu-cen**)

Cy5-5'-ATGAATCTGGACTTC (SEQ ID NO: 26)

### Cy3 labelled oligonucleotides

**E** (**Cy3-wt-ter2**)

10 Cy3-5'-GGTTCCTTAGTGTTAG (SEQ ID NO: 34)

**F** (**Cy3-mu-ter2**)

Cy3-5'-TGTTCCTTAGTGTTAG (SEQ ID NO: 35)

**G** (**Cy3-wt-ter**)

Cy3-5'-CGGACTTCATTTTTC (SEQ ID NO: 36)

15 **H** (**Cy3-mu-ter**)

Cy3-5'-TGGACTTCATTTTTC (SEQ ID NO: 37)

### Rhodamine labelled oligonucleotides

**I** (**Rho-wt-ter**)

20 Rho-5'-CGGACTTCATTTTTC (SEQ ID NO: 38)

**J** (**Rho-mu-ter**)

Rho-5'-TGGACTTCATTTTTC (SEQ ID NO: 39)

**K** (**Rho-wt-cen**)

Rho-5'-ATGAATCCGGACTTC (SEQ ID NO: 40)

25 **L** (**Rho-mu-cen**)

Rho-5'-ATGAATCTGGACTTC (SEQ ID NO: 41)

### Hybridisation conditions

30 DNA target and fluorophore labelled oligonucleotides (usually 0.5  $\mu$ M) were kept at the hybridisation temperature in a commercially available hybridisation buffer containing 1 X PCR buffer - 10 mM Tris-HCl pH 8.3, 50 mM KCl (Perkin Elmer, catalog # N808-0010).

### Spectrofluorometry

35 Equal volumes of target DNA and fluorophore labelled oligonucleotide were mixed in a tube and kept at the hybridisation temperature for 5 min until reading of the spectrum. The final concentrations of all oligonucleotides were 0.25  $\mu$ M, unless otherwise stated. Spectra were obtained using a Perkin Elmer LS50

Spectrofluorometer. The cuvette holder and the plastic cuvette (VWR (Merck) catalog # 122455/100) was kept at a constant temperature (23°C, 30°C or 45°C) by connection to a thermostated water bath. Individual oligonucleotide mixtures were kept at the hybridisation temperatures in a thermostated water bath. Spectral ranges are provided in Table 5. The excitation spectrum was obtained by keeping the  $E_{m_{max}}$  constant, whereas the emission spectrum was obtained by keeping the excitation wavelength at  $E_{x_{max}}$ .

Dye-oligo	$E_{x_{max}}$ nm	$E_{m_{max}}$ nm	Ex spectrum nm	Em spectrum nm
Cy3-oligo	550	570	500 – 800	500 – 800
Cy5-oligo	649	670	550 – 800	550 – 800
Rhodamine-oligo	570	587	500 – 800	500 – 800

**Table 5.** Fluorescence properties of the three fluorophores used.

### **Experiments**

#### **Experiment 1.**

Two pairs of Cy5-5' labelled oligoes which were either perfect match or mismatch, respectively, in the 5'-ends or in the middle positions, relative to the target were hybridised to a wildtype target (W1). Spectra were obtained at RT (23°C).

#### **Experiment 2.**

Same as Experiment 1, but in addition a set of rhodamine labelled oligonucleotides were used. Spectra were obtained at 30 °C after 5 min of hybridisation.

#### **Experiment 3.**

A Cy5 labelled oligonucleotide (complementary to the wildtype target) was hybridised to the wildtype target (W2), the mutant target (M) or the dummy target (Z), respectively. Spectra were obtained at 30 °C after 5 min of hybridisation.

#### **Experiment 4.**

Hybridisation of the Cy5 labelled wildtype oligonucleotide to either the wildtype target (W2), the mutant target (M) or both W2 and M simultaneously.

**Experiment 5.**

Hybridisation of two sets of Cy3 and Cy5 labelled oligonucleotides simultaneously (only partial measurements).

5 **Experiment 6.**

Repetition of experiment 3 and 4.

**Experiment 7.**

Hybridisation of the Cy5 labelled wildtype oligo nucleotide to dsDNA.

10

**Experiment 8.**

Hybridisation of a Cy5 labelled oligonucleotide to a polymorphic site in the muscle glycogen synthase promoter.

15 **Experiment 9.**

Cy5 labelled oligonucleotides with LNA monomers are hybridised to a polymorphic site in the muscle glycogen synthase promoter.

**Experiment 10.**

20 Experiment 8 is repeated but a PCR-product from genomic DNA is used as target

Experiment 1		Experiment 2		Experiment 3		Experiment 4		Experiment 5		
A	Terminally, Cy5	A	Terminally, Cy5	A	Terminally, Cy5	A	Terminally, Cy5	F	Terminally, Cy3	
A-W1		A-W1		A-W2		W2		F-W3		
B		B		A-M		M		F-A		Terminally, Cy3/Cy5
B-W1		B-W1		A-Z		A-W2		F-A-W3		
C	Internally, Cy5	C	Internally, Cy5			A-M	1=0,25 2=0,125	F-B-W3	Terminally, cy5	
C-W1		C-W1				A-W2-M-1		A		
D		D				A-W2-M-2		A-W3		
D-W1		D-W1						B		
W1	Controls	I	Terminally, Rhodamine					B-W3	Control	
Buf		I-W1						Buf		
		J	Internally, Rhodamine							
		J-W1								
		K								
		K-W1								
		L								
		L-W1								

Table 6 summarises the oligonucleotide compositions used in each of the of 5 experiments. The nomenclature of the oligonucleotides is deciphered above. Terminally implies that the mutation is addressed by the nucleotide, which is placed at the 5'-end attached to the fluorophore via a linker. Internally means that the mutation is addressed by an unlabelled nucleotide placed in the centre position of the oligonucleotide.

Spectral data were analysed by PCA using the Unscrambler program version 7.6 from Camo AS, Olav Tryggvasons 24, N-7011 Trondheim, Norway.

## RESULTS AND DISCUSSION

The major results are presented below. Figure 4 gives an overview of the target wildtype and mutant DNA and of all (partially) complementary labelled oligonucleotides used in the present examples.

### Experiment 1

Two pairs of Cy5-5' labelled oligonucleotides which were either perfect match or mismatch, respectively, at the 5'-ends or at the middle position, relative to the target were hybridised to a wildtype target (W1). Spectra were obtained at RT (23 °C).

Figure 5 shows, that when the 4 different oligonucleotides (A, B, C and D) are hybridised to a target DNA (W1) it changes their placement in the scores plot - i.e. the hybridisation changes the fluorescence spectra.

The first prefix (1 or 2) denotes the replicate number, whereas the second prefix only represent two fluorescence measurements on that same replicate solution. This implies, that samples with the prefix 1.1 and 1.2 and the prefix 2.1 and 2.2 respectively should be placed closer to each other in the plot – only showing the difference in fluorescence spectra obtained successively in time. The plot shows that this is the case. There is most often a larger variation within replicates than within two fluorescence measurements on the same replicate. In conclusion there are some differences between replicates and double measurements. It should be noted that the hybridisations reactions were not allowed to hybridise for exactly the same time in this first shot experiment.



**Conclusion of experiment 1:**

Hybridisation changes the fluorescence emission spectra.

The 4 hybridised oligonucleotide combinations (A-W1, B-W1, C-W1 and D-W1)  
5 results in different emission fluorescence spectra.

**Experiment 2**

Two pairs of Cy5-5' labelled oligonucleotides which were either perfect match or  
10 mismatch, respectively, in the 5'-ends or in the middle positions, relative to the target were hybridised to a wildtype target (W1). Spectra were obtained at 30 °C.

Figure 6 shows, that when the 4 different oligonucleotides (A, B, C and D) are  
15 hybridised to a target DNA (W1) it changes their placement in the scores plot - i.e. the hybridisation changes the obtained fluorescence spectrum (emission).

For all the 4 oligonucleotides the hybridisation to the target DNA moves the samples  
more to the right in the scores plots, which implies that the score value of PC1  
increases.

20 In Figure 7 it is seen, that the 4 oligonucleotides (A, B, C and D) hybridised to the target DNA (W1) is placed in each group of the scores plot - i.e. the 4 hybridised oligonucleotides gives different fluorescence spectra (emission). However, the variation within each group is just a little smaller than the variation between groups.

25 On Figure 8 it is observed, that when the 4 different oligonucleotides (A, B, C and D) are hybridised to a target DNA (W1) it changes their placement in the scores plot - i.e. the hybridisation changes the obtained fluorescence spectrum (excitation).

30 On Figure 9 it is observed, that the 4 oligonucleotides (A, B, C and D) hybridised to the target DNA (W1) is placed in each quadrant of the scores plot - i.e. the 4 hybridised oligonucleotides obtain different fluorescence spectra. The excitation spectra shows much smaller variation within each group compared to the variation between the groups. Thus, the excitation spectrum seems to provide better data.

35 Conclusion of experiment 2:

Hybridisation changes both the fluorescence emission and the excitation spectra.

The 4 hybridised oligonucleotides (A-W1, B-W1, C-W1 and D-W1) gives different fluorescence spectra (both emission and in excitation).

5 The differences were most prominent for the excitation spectra.

### Experiment 3

10 A Cy5 labelled oligonucleotide (complementary to the wildtype target) was hybridised to the wildtype target (W2), the mutant target (M) or the dummy target (Z), respectively. Spectra were obtained at 30 °C after 5 min of hybridisation.

15 Figure 10 (excitation spectrum) shows, that when oligonucleotide A are hybridised to a target DNA (W2 or M) it changes its placement in the scores plot - i.e. the hybridisation changes the obtained fluorescence spectrum (excitation). Note that the non-specific target (Z) groups near to A alone.

20 It is also noted, that the 3 hybridised oligonucleotides (A-W2, A-M and A-Z) is placed in each group in the scores plot - i.e. the 3 hybridised oligonucleotides may represent different fluorescence spectra.

In Figure 11 (emission spectrum) the same tendency as in Figure 10 is seen, but the results are based on the emission spectra and thus shows a larger variation within each group as compared to the variation between the groups.

25 Conclusions of experiment 3:

Hybridisation changes the fluorescence spectrum. (both the excitation and the emission spectra).

30 The 3 hybridised oligonucleotides (A-W2, A-M and A-Z) separates out in different groups.

### Experiment 4

35 Hybridisation of the Cy5 labelled wildtype oligonucleotide to either the wildtype target (W2), the mutant target (M) or both W2 and M simultaneously.

Figure 12 (excitation spectrum) shows that when oligonucleotide A is hybridised to a target DNA (W2, M or W2/M) it changes its placement on the scores plot - i.e. the hybridisation changes the obtained fluorescence spectrum (excitation).

5

It is observed, that the 2 hybridised oligonucleotides A-W2 and A-M are placed in each group in the scores plot. The hybridised oligonucleotides A-W2-M (W2 and M targets are present simultaneously) are placed in between the A-W2 group and the A-M group. No clear difference between the two W2-M concentrations (corresponding to the prefix -1 or -2 in the A-W2-M samples) was observed, i.e. the two concentrations (0,25 or 0,125  $\mu$ M of each oligo nucleotide) did not influence on the grouping.

10

Figure 13 (emission spectrum) gives the same picture as is depicted in Figure 12. Compared to the results observed in Figure 12 which is based on excitation spectra, the results based on the emission spectra indicates a smaller variation within each group in relation to the variation between the groups.

15

The conclusions of experiment 4 are:

20

Hybridisation changes the fluorescence spectrum. (both in excitation and in emission).

The 2 hybridised oligonucleotides A-W2 (wildtype homozygote) and A-M (mutant homozygote) gives different fluorescence spectra (both in excitation and in emission) – and the A-W2-M (heterozygote) is grouped in between.

25

No clear difference between the W2 and M concentrations (for the A-W2-M samples) is observed.

### Experiment 5

30

Two sets of oligonucleotides, a Cy3 labelled and a Cy5 labelled covering two nucleotide regions of the target W3 were hybridised alone or simultaneously to the DNA-target W3 and scanned at Cy3 settings or at Cy5 settings. Unfortunately the Cy3 spectrum were recorded under saturating conditions for most of the wavelengths. Therefore we present only data of the Cy5 spectra.

35

As expected A and A-W3 and B and B-W3 groups in separate groups (Fig. 14) in the excitation spectrum. F-A and F-A-W3 lie very close to each other but it may be possible to separate them in two subgroups (Fig. 14, upper left, note the scale on PC1).

5

An examination of the emission spectrum (Fig. 15) supports these patterns. A-W3 and F-A-W3 lie within the same group i.e. the Cy5 spectrum is relatively unaffected by the presence of the Cy3 labelled oligonucleotide. This is supported by the observation that A and A-F lie within the same group. Note that A-F is distinct from A-F-W3 as opposed to the situation on the excitation spectrum (Fig. 14) where they were grouped.

10

Conclusions of experiment 5 are:

15

Hybridisation changes the fluorescence spectrum. (both in emission and in excitation).

The combinations A and A-W3, and B and B-W3 separates as expected and are clearly distinguished from each other.

20

In the presence of F (the Cy3 labelled oligonucleotide complementary to the utmost upstream position of the target W3) F-A and F-A-W3 are grouping close together (excitation), but are separated in the (emission) PCA plot.

There is a clear difference between F-A-W3 and F-B-W3 which means that in the presence of a second probe (the Cy3 labelled F) the full match (A-W3) and the mismatch (B-W3) can still be differentiated.

25

### Experiment 6

30

This experiment was performed one year after the experiments 3 and 4. New batches of reagents including oligoes were used and all technical staff was new to the technique. Data analysis was done by the same person.

The experiment reproduced the results previously obtained in experiments 3 and 4 (not shown).

Conclusions of experiment 6 are:

35

The principle of the invention is confirmed.

The technique is stable over time, changes of reagents and independent of specific technical skills of the staff. The major impact of letting the same person perform the data analysis was to reduce the time used to a few minutes.

5

### Experiment 7

dsDNA was created by hybridising equimolar amounts of oligonucleotide W2 to its complementary strand R (Figure 4) mixing 1  $\mu$ M solutions of the two strands in standard buffer and incubated for 10 minutes at 30°C. The dsDNA was denatured by heating to 94 °C for 10 min and cooled on ice. Equimolar amounts of Cy5-labelled probe A was added, and the reaction mixture was incubated for 5 minutes at 30°C The spectra and data analysis was performed as previously described.

The results are shown in Figure 19. The hybridisation mixture groups the wildtype dsDNA-probe complex (AW2-R) and the mutant dsDNA-probe complex (AM-Rm) in two clearly separated areas of the score plot. AW2-R- denotes dsDNA not denatured before the reaction with the probe. Hybridisation to single stranded DNA (AM and AW2) and non-specific single stranded DNA (AZ) is included for illustrative purposes. These latter complexes are not normally present in the reaction vial. The most consistent grouping is seen with dsDNA targets denatured before annealing to the labelled probe.

Conclusions of experiment 7 are:

25

Hybridisations with dsDNA as target changes the fluorescence spectra. The wildtype and mutant dsDNA complexes grouped in separate groups, suggesting that the Cy5-labelled probe displaced the complementary strand, i.e. no elimination of the complementary strand was necessary. This means that the reaction can be performed as a homogenous reaction with no purification steps needed.

30

### Experiment 8

The experiment was repeated with a newly detected polymorphism in the promoter of human muscle glycogen synthase named Poly 2 related to diabetes mellitus (pat-

35

ent pending). Sequences of oligonucleotides were the following:

WA Poly2 antisense wildtype (SEQ ID NO: 61):

5'-atggggccagacccgagattctgggatcccagccccGtccccgcctcagatccagaagtcagc

5

WS Poly2 sense wildtype (SEQ ID NO: 62):

3' -taccocggtctgggctctaagaccctagggtcgggggaggggcggagtctaggtcttcaggtcg

MA Poly2 antisense mutant (SEQ ID NO: 63):

10

5' -atggggccagacccgagattctgggatcccagccccGtccccgcctcagatccagaagtcagc

MS Poly2 sense mutant (SEQ ID NO: 64):

3' -taccocggtctgggctctaagaccctagggtcggggcaggggcggagtctaggtcttcaggtcg

15

WP Poly2 wild type probe (SEQ ID NO: 65): Cy5-5'-Gggggctgggat

dsDNA was created as described in above. Reaction conditions was exactly as described in the previous experiments. Figure 20 shows the score plot (excitation) from the reaction mixtures. The complex between the probe and the target was grouped in separate areas for the wild type and the mutant.

20

Conclusions of experiment 8 are:

Different spectral changes was obtained for the wildtype and the mutant of Poly2 in human muscle glycogen synthase. The invention does not succumb to the structural implications of oligonucleotides with a high GC-content.

25

### Experiment 9

Experiment 8 is repeated but the Cy5-labelled probe A is substituted with the A-LNA

probe: C<sup>L</sup>T<sup>L</sup>T<sup>L</sup>T<sup>L</sup>TA<sup>L</sup>CT<sup>L</sup>TC<sup>L</sup>AG<sup>L</sup>GC-5' cy5-wt-ter (SEQ ID NO:58)

wherein L = LNA monomer

35

### Experiment 10

DNA is isolated from blood using a commercial method from Roche. The amplification reaction (PCR) in 25  $\mu$ L consists of 100 ng of target, 1  $\mu$ M of primer 3611s and 3611as each, 0.5 unit Taq polymerase (Quiagen Core), 0.2 mM of each of the four dNTP, 1.5 mM  $MgCl_2$ , 5  $\mu$ L Q-solution (from Quiagen), and PCR-buffer accompanying the Taq polymerase. The PCR-reaction is performed with an initial denaturation at 94°C for 5 minutes, 35 cycles of denaturation at 94°C for 40 seconds, annealing at 56°C for 40 seconds and extension at 72°C for 40 seconds. The final extension step is extended to 10 minutes. The reaction solution is heated to 94°C for 5 minutes, cooled on ice, and the Cy5 labelled probe A is added to a final concentration of 1  $\mu$ M. The solution is diluted by PCR-buffer to 500  $\mu$ L and incubated at 30°C for 5 minutes and the spectra was recorded and data processed as previously.

#### Primers

3611s	AGAACATACAAGCAAAGCCA	(SEQ ID NO: 59)
3611as	GAGGAACCTTAGGTGTCCTTC	(SEQ ID NO: 60)

#### Reference List

- Bentzen, J., T. Jorgensen, and M. Fenger. "The effect of six polymorphisms in the apolipoprotein B gene on parameters of lipid metabolism in a Danish population." Clin.Genet. In press (2002).
- Bentzen, J. et al. "The influence of the polymorphism in apolipoprotein B codon 2488 on insulin and lipid levels in a Danish twin population." Diabet.Med. 19.1 (2002): 12-18.
- Fenger, M. et al. "Impact of the Xba1-polymorphism of the human muscle glycogen synthase gene on parameters of the insulin resistance syndrome in a Danish twin population [In Process Citation]." Diabet.Med. 17.10 (2000): 735-40.
- Jarvik, G. P. "Genetic predictors of common disease: apolipoprotein E genotype as a paradigm." Ann.Epidemiol. 7.5 (1997): 357-62.

Peltomaki, P. "Deficient DNA mismatch repair: a common etiologic factor for colon cancer." Hum.Mol.Genet. 10.7 (2001): 735-40.

5 Yasutomo, K. et al. "Mutation of DNASE1 in people with systemic lupus erythematosus." Nat.Genet. 28.4 (2001): 313-14.

### Sequence table

10 Highlighted residues correspond to the position of polymorphisms in the target nucleic acid sequence.

*Apolipoprotein B mutations related to atherosclerosis* (Bentzen et al. 12-18; Bentzen, Jorgensen, and Fenger)

Apolipoprotein B 3500 mutation:

15 SEQ ID NO 1 5'-AGCACACGGTCTTC (wt)

SEQ ID NO 2 5'-AGCACACAGTCTTC (mut)

Apolipoprotein B 2488 polymorphism:

SEQ ID NO 3 5'-CGAGAGACCCTAGAAGA (allele 1)

SEQ ID NO 4 5'-CGAGAGACTCTAGAAGA (allele 2)

20

*Apolipoprotein E polymorphism (apoE2, E3 and E4) related to neurological diseases* (Jarvik 357-62)

SEQ ID NO 5 5'-GACGTGTGCGGCCGC apoE codon112-cys

SEQ ID NO 6 5'-GACGTGCGCGGCCGC apoE codon112-arg

25 SEQ ID NO 7 5'-CAGAAGTGCCTGGCA apoE codon 158-cys

SEQ ID NO 8 5'-CAGAAGCGCCTGGCA apoE codon 158-arg

*Human muscle glycogen synthase polymorphism related to diabetes mellitus* (Fenger et al. 735-40)

30 SEQ ID NO 9 5'-ACTCCATTCTAGAGT

SEQ ID NO 10 5'-ACTCCATCCTAGAGT

*Dnase1 mutations related to rheumatological diseases* (Yasutomo et al. 313-14)

SEQ ID NO 11 5'-GGGGCATGAAGCTGCT (wt)

35 SEQ ID NO 12 5'-GGGGCATGTAGCTGCT (mutation)



*methylene tetrahydrofolate reductase polymorphisms related to osteoporosis*

SEQ ID NO 13 5'-TGCGGATCGATTTC (wt)

SEQ ID NO 14 5'-TGCGGACCGATTTC (mutation)

5

*Mismatch repair gene mutations related to cancer (Peltomaki 735-40)*

SEQ ID NO 15 5'-GAAGAAGGCT (wt)

SEQ ID NO 16 5'-GAAGGCGGCT (mutation)

10

*General and specific 16S-RNA's related to infectious diseases*

SEQ ID NO 17 5'-AGGAGGTGATCCAACCGCA (general capture-probe for 16S)

SEQ ID NO 18 5'-GGCGCTTACCACTTTGTGATTCAT (specific capture-probe for Enterobacteriaceae 16S)

SEQ ID NO 19 5'-GGAAGAAGCTTGCTTCTTTGCTGAC (specific capture-probe for

15

E. Coli-ECA75F 16S)

*Experimental oligoes and probes (the oligonucleotides were selected from a region of the apolipoprotein B gene at the codon3611 polymorphism):*

SEQ ID NO 20

20

W1 5-ctaagaaccagaagatcagatggaaaaatgaagtccggattcattctgggtctttccagagccaggtcga

SEQ ID NO 21

W2 5-accagaagatcagatggaaaaatgaagtccggattcattctgggtctttc

SEQ ID NO 22

M 5-ACCAGAAGATCAGATGGAAAAATGAAGTCCAGATTTCATTCTGGGTCTTTC'

25

SEQ ID NO 23

A CTTTTTACTTCAGGC-5' (Cy5-wt-ter)

SEQ ID NO 24

B CTTTTTACTTCAGGT-5' (Cy5-mut-ter)

SEQ ID NO 25

30

C CTTTCAGGCCTAAGTA-5' (Cy5-wt-cen)

SEQ ID NO 26

D CTTTCAGGTCTAAGTA-5' (Cy5-mut-cen)

SEQ ID NO 27 5'-AACACCCAGGATCCT (BRCA1 gene codon 1313 wt)  
 SEQ ID NO 28 5'-AACACCTAGGATCCT (BRCA1 gene codon 1313 mut)  
 SEQ ID NO 29 5'-CTG GAACAGTCTGGG (BRCA1 gene codon 1541 wt)  
 SEQ ID NO 30 5'-CTG GAATAGTCTGGG (BRCA1 gene codon 1541 mut)

5

In the following the possible variation in probes is described with reference to the ApoE polymorphism (SEQ ID No 5-8). *Apolipoprotein E polymorphism (apoE2, E3 and E4) related to neurological diseases* (Jarvik 357-62)

10 The residue corresponding to the polymorphism is located in the 3' end of the probe. The label may be linked to the 3' or the 5' terminal residue.

SEQ ID NO 42: 5'-ATGGAGGACGTGT apoE codon112-cys

SEQ ID NO 43: 5'-ATGGAGGACGTGC apoE codon112-arg

15 SEQ ID NO 44: 5'-GACCTGCAGAAGT apoE codon 158-cys

SEQ ID NO 45: 5'-GACCTGCAGAAGC apoE codon 158-arg

The label is located at the position of the polymorphism.

20 SEQ ID NO 46: 5'-CATGGAGGACGTG-label apoE codon112-cys/arg

SEQ ID NO 47: 5'-TGACCTGCAGAAG-label apoE codon 158-cys/arg

The residue corresponding to the polymorphism is located in the 3' end of the probe. These probes will hybridise to the opposite string compared to X1-X4:

25

SEQ ID NO 48: 5'-CCAGGCGGCCGCA apoE codon112-cys

SEQ ID NO 49: 5'-CCAGGCGGCCGCG apoE codon112-arg

SEQ ID NO 50: 5'-ACACTGCCAGGCA apoE codon158-cys

SEQ ID NO 51: 5'-ACACTGCCAGGCG apoE codon158-arg

30

Probes with the label in the position of the polymorphism. Hybridise to the opposite string as compared to Y1-Y2.

SEQ ID NO 52: 5'-ACCAGGCGGCCGC-label apoE codon112-cys/arg

SEQ ID NO 53: 5'-TACACTGCCAGGC-label apoE codon158-cys/arg

35

Probes with the residue complementary to the polymorphism in a central position. The label may be in the 3' or 5' end. These probes hybridise to the opposite string compared to SEQ ID NO 5 to 8.

- 5      SEQ ID NO 54: 5'- GCGGCCGCACACGTC apoE codon112-cys  
SEQ ID NO 55: 5'- GCGGCCGCGCACGTC apoE codon112-arg  
SEQ ID NO 56: 5'- TGCCAGGCACTTCTG apoE codon 158-cys  
SEQ ID NO 57: 5'- TGCCAGGCGCTTCTG apoE codon 158-arg